COMMONWEALTH OF AUSTRALIA

Copyright Regulations 1969

Warning

Chapter 10

Evidence-based midwifery: finding, appraising and applying evidence in practice

Lesley Page, Michael Corkett, Rona McCandlish

In the first edition of this book Lesley Page (Page 2000) described the experiences which began her journey of development as an evidence-based midwife saying:

> My own interest in what we now call evidence-based care started when I practised as a midwife in Canada, providing care for women and their families through the whole process of pregnancy and birth, for the first time in my experience as a midwife. As I got to know these women as individuals, I became increasingly aware of the importance of doing the right things for them as individuals. Inevitably, as we came to know each other through the course of pregnancy, the parents would start to ask why certain things were undertaken as a routine. Intuitively, I guessed that many of these routines, which were imposed in our large maternity hospital, were unfounded. It was only when I started to search for evidence, so that I could make an argument for abandoning some routines for the women in my care, that I began to realize just how senseless some of the routines were. For example, there was a very strict rule that there was to be absolutely nothing to drink or eat in labour, yet many of the women I cared for wanted the freedom to eat and drink in labour. Thus, I started to investigate the evidence on the topic. This took me many hours. I contacted others who were undertaking research in the area in question, and used the library. Now, with a number of sources of synthesized evidence, the search for evidence is easier in some areas.

As reflected in Page's words evidence-based practice requires an awareness that care affects a constellation of important outcomes, ranging from the physical and emotional, personal and family integrity, to the wider social and economic. Finding and understanding evidence to inform decisions and choices about care will mean confronting uncertainty about what is best, or better, and dealing with frustration when the usefulness of evidence may not be straightforward. Although measurements of death or serious morbidity are conventionally used as key indicators of the effectiveness and quality of care these do not encompass the reality that pregnancy, birth and family life are more than physical 'events'. Measurement of the impact of maternity care is usually more complicated than simply 'counting' outcomes.

This throws up the most challenging aspect of practising evidence-based midwifery: the need to weigh up the validity and applicability of evidence about potential benefits and risks of certain choices and decisions whilst maintaining open and honest communication with women, families and colleagues. As a consequence evidence-based practice means that midwives often find themselves questioning long-standing routines in systems that are not easy to challenge. Effective questioning demands accumulating convincing evidence about what is likely to be effective care, developing effective skills to assemble and evaluate evidence, and then using it to change practice for the better. Simply retrieving evidence is not

sufficient to practising evidence-based care; there is no point in collecting highest quality best evidence if it is not used to improve practice and outcomes of care. Therefore the approach we describe in this chapter makes responding to the needs of women and their families the starting point for asking questions.

There are two fundamental questions in evidence-based midwifery:

1. Is what I intend to do likely to do more good than harm?
2. Am I spending my time doing the right things?

Every midwife can develop key skills to ask these questions, work through the answers and apply new learning, knowledge and insights effectively in her or his practice. The skills of lifelong learning and the ability to undertake independent enquiry are therefore crucial and this has implications both for the way in which midwives learn in basic education programmes and for continuing professional education.

In this chapter we will discuss evaluating care and effective care, and briefly outline some of the influences to the development of evidence-based maternity care in the UK.

EVALUATING AND ASSESSING THE EFFECTIVENESS OF CARE

During the second half of the 20th century in the UK the drive to improve safety of birth through hospitalization and the use of technology and medical diagnosis was marked by imposition of treatments and care that were largely unevaluated. This was demonstrated for example, by routine use of perineal shaving and enemas when women were admitted to the labour ward, and more recently in the increasing proportion of women who experience induction of labour and routine use of electronic fetal monitoring. Publication of the book Effective Care in Pregnancy and Childbirth (Chalmers et al 1989) brought about a greater awareness that the effects of maternity care should be rigorously evaluated. This seminal work was an important foundation for practitioners and consumers accessing care to understand different forms of evidence, and ways of judging the strength of that evidence, in deciding the probable effects of care. In addition, it synthesized much of the relevant evidence for reference. From this, the Cochrane Collaboration and the Cochrane Library were developed (www.cochrane.org).

The ability to evaluate and understand effects of care is crucial to the practice and provision of ethical healthcare. At its most benign not knowing the effects of an inert intervention mean its continued use with consequent misuse of limited healthcare resources. At worst, lack of evidence of effectiveness means that harmful practices are applied. There are clear instances of this in pregnancy and childbirth care when there has been unquestioned adoption of innovation which has led to avoidable tragedies. For example, Silverman (1980) describes the story of the 'epidemic' of neonatal retrolental fibroplasia which took hold in the 1950s in the USA. This is a dramatic example of a well-intentioned treatment – in this case the administration of high concentrations of oxygen to premature babies – which had unknown and unsuspected adverse consequences. The oxygen regimens resulted in blindness in large numbers of the babies who had received the treatment. This unintended harm could have been minimized if practitioners who advocated the use of high-concentration oxygen had been committed to asking questions about effectiveness (Box 10.1). This would have allowed them to evaluate whether the treatment they believed would work actually caused more good than harm. Instead, extremely vulnerable babies were exposed to a regimen of oxygen therapy which caused serious life-long morbidity. For lack of rigorous evaluation this form of care was continued for a decade longer than it should have been, and the concern it caused

Box 10.1

Effectiveness: . . . a measure of the extent to which a specific intervention, procedure, regimen, or service, when deployed in the field in routine circumstances, does what it is intended to do for a specific population.

Cochrane 1972

still acts as an impediment to carrying out research about different oxygen regimes for neonates (Silverman 1998).

One of the difficulties about the term 'effectiveness' is that it is understood differently by different people. For example, imagine a policy of routine augmentation of labour is introduced for a trial period in your local labour ward to reduce the length of time women spend in labour care and that the effectiveness of the policy was measured after six months. The results of the evaluation show a reduction in the time women spent in labour care and therefore it is proposed that the policy be adopted. However, the midwife responsible for collating unit statistics points out that compared to the 6 month period previous to the introduction of the new policy, there has been a high rate of analgesia use and a high proportion of women have said they were dissatisfied with their experience of labour. This would suggest that measuring effectiveness using length of labour alone is inadequate and that adopting the policy would be unjustified as it could be associated with more harm than good. As this example demonstrates, judgements about the effects of care during pregnancy and childbirth, as in other areas, are neither value free nor situation free. Different observers see different problems and often reach different conclusions (Susser 1984, in Chalmers et al 1989) and it is vital to take these factors into account.

EVIDENCE-BASED CARE

The movement towards 'evidence-based healthcare' (Gray 1997) has been important in helping healthcare professionals and others, including policy-makers and managers, understand that research is carried out so that the results it yields can be used. It challenges practitioners to consider whether they are simply practising in the ways they were first taught, or in response to being 'told' by someone in authority, or perhaps from decisions based on personal opinion. Evidence-based practice requires active searching for, and appraisal of, research evidence to inform decisions about tests, treatments, patterns of practice, and policy. Although personal experience is an important basis to understanding what works and why, it is rarely wide enough to give objective answers about the effects of particular tests and treatments.

Gray (1997, p. 213) describes evidence-based clinical practice as 'the judicious use of the best evidence available so that the clinician and the patient arrive at the best decision, taking into account the needs and values of the individual patient'. Evidence should be used to inform decisions in a number of areas: policy, guidelines for practice, the appropriate organization of care, public health decisions (about the use of resources, for example), clinical decisions and information to help women's choice, health promotion and education for parenting.

Other chapters describe ways of using evidence in practice and described the sources of information used to inform decisions about care. These are:

- individual values or preferences;
- the clinical examination;
- research evidence;
- the context of care.

In that chapter, five steps for the use of evidence in practice are described:

1. finding out what is important to the woman and her family;
2. using information from the clinical examination;
3. seeking and assessing evidence to inform decisions;
4. talking it through;
5. reflecting on outcomes, feelings and consequences.

We now focus on the third step: finding and critiquing the evidence.

Although evidence-based midwifery is about using rather than doing research, in common with primary research avoiding bias is fundamental when seeking, selecting and assessing evidence.

The process of using evidence in practice includes:

- framing clear and relevant questions that will lead to an effective search;
- planning an efficient search to answer the question;
- assessing and weighing up the evidence.

FRAMING CLEAR QUESTIONS

PRINCIPLES TO GOOD QUESTIONS

A pre requisite to finding the right and best evidence is the ability to convert a precise, yet possibly vaguely expressed, need into an answerable, focused, structured question (Rosenberg & Donald 1995, Straus et al 2005). Well-structured questions have four key components:

- Population – in the case of midwifery, women.
- Intervention – cause, prognostic factor, treatment.
- Comparison – control or comparative intervention.
- Outcome – ways in which the intervention effect is measured.

These are mnemonically referred to as PICO and provide a robust framework by which to execute a search. An additional component relating to study design can also be employed, and prove particularly helpful when improving the efficiency of a search (see Planning an efficient search). The most useful types of study design that may form part of a search include:

- Systematic reviews – literature reviews focused on a single question which identify, appraise and synthesize all relevant high quality research evidence.
- Meta-analyses – systematic reviews or overviews which employ quantitative methods to summarize the results.
- Randomized controlled trials – patients are randomized into intervention and control groups, and followed up for outcomes.
- Cohort studies – identification of two groups (cohorts), one of which normally received exposure to an intervention and one which did not, and followed up for outcomes.
- Case-series – reports on a series of patients with outcomes.

When including study design in a search it is important to remain aware of potential compromises to quality. Always consider the quality of study design – the number of patients treated, or group size; the duration of the study; the objective measurement of outcomes; and, the elimination of bias.

To assist the understanding and application of PICO to the midwifery setting Table 10.1 presents a modification of the work of Sackett et al (1997) and Straus et al (2005, p. 257–259).

It is worth noting that in the maternity services the majority of women and families start off being healthy, and the aim should be to keep them that way. Because of this, the potential for doing harm is greater. One of the problems of maternity care in much of the industrialized world is the routine treatment of women as if they were a high-risk population with a high probability of adverse outcomes. In reality, the risk of an adverse outcome is lower than it has ever been. Therefore, many of the questions that midwives ask arise from the need to determine whether or not a woman is in a high-risk group; or, has the woman a higher chance than usual of an adverse outcome.

EXERCISE

Consider your own practice and think of a number of questions about the routines you undertake, then apply the PICO framework to each.

For example, you may be looking after a woman who is experiencing leg cramps. The question might be:

- In a woman with a normal pregnancy, who suffers severe leg cramps at night, is calcium supplementation likely to help; and are there any likely harmful side effects to this supplementation?

Following the PICO framework the question can be illustrated as follows:

- Population = women with a normal pregnancy, who suffer severe leg cramps at night.
- Intervention = calcium supplementation.
- Comparison = an alternative intervention, for example exercise.
- Outcome(s) = reduction of leg cramps; harmful side effects to supplementation.

Now consider the following situation:

Mrs Smith is approaching term. She has prepared a birth plan indicating that she wishes to avoid intervention in labour. At her 39 weeks visit, the fundal height is 39 cm, the fetus is active, her blood pressure is 120/85, and there is no protein

Table 10.1 Comparisons of good questions. (Adapted from Sackett et al 1997, p. 27 with kind permission)

1. Woman or problem	2. Intervention (cause, prognostic factor, treatment)	3. Comparison intervention (if necessary)	4. Outcome(s)
Tips for building			
How would I best describe a group of women similar to mine?	Which main intervention or complication or 'risk factor' am I considering?	What is the probability of adverse outcome?	What can I hope to accomplish? What else would be affected?
Examples			
In women in early pregnancy who are vomiting most of the day.	Acupressure	No Acupressure	Acupressure leads to a reduction of vomiting and the experience of nausea?
In women without other complications	Who are grand multiparous (greater than gravida 5)	When compared with women who are less than gravida 5	Is there a greater probability of excessive bleeding, a need for blood transfusion, illness or death?
In women of 26 years of age	Who have an amniocentesis for the diagnosis of Down's syndrome	Who have not had amniocentesis for diagnosis of Down's syndrome	What is the probability of miscarriage? What is the probability of Down's syndrome? What are the sensitivity and specificity of the test?
In nulliparous women without complications	Who have an elective prelabour caesarean section	Rather than allowing labour and vaginal birth	What will the effect on perinatal mortality and morbidity, and maternal mortality and morbidity be?
In pregnant women	Who are over 40 years old	Compared with women of under 40 years of age	Is there a greater probability of adverse outcomes (e.g. perinatal mortality and higher intervention rates) as a result of age alone?

in the urine. At this visit, Mrs Smith wonders whether it is enough to listen to the fetal heart regularly or whether she should think about having continuous electronic monitoring.

Using what you have learnt about the PICO framework, and the examples provided in Table 10.1, write a question to guide your search for evidence in the box opposite. (Box 10.2)
The question might read:

Box 10.2 Write your response Here

In women at term, with no problems during pregnancy, would continuous electronic fetal monitoring, rather than intermittently listening to the fetal heart, lead to lower mortality and morbidity, and how would it affect intervention rates?

OTHER TYPES OF QUESTION

The questions considered so far have been concerned with the effect of particular interventions on outcomes of care. While this represents an important part of clinical practice, a further large and important field for consideration encompasses decisions made concerning risk, or the probability of adverse outcomes.

The most effective way to answer questions in this realm is to use systematic reviews and/or meta-analyses, which combine the results of a number of studies. This approach is particularly beneficial in those areas where a considerable number of individual studies exist, and it would be beyond the available time of clinicians to read and appraise each. For example, when Olsen undertook his meta-analysis of the studies of home birth, he identified 65 separate studies that met his criteria (Olsen 1997).

Midwives are fortunate to have available to them several evidence-based medicine resources that bypass the need to read and appraise numerous individual studies. These include the National Guideline Clearing House, a repository of high-quality international clinical guidelines; clinical guidelines produced in the UK by the National Institute for Health and Clinical Excellence (NICE), in particular those concerned with women's and children's health; the Cochrane Library, an extensive collection of rigorously appraised systematic reviews; and MIDIRS, a specialist midwifery information service. Further details on these resources can be found in Table 10.2.

Table 10.2 Resources for searching for evidence

Resource	Description	Access	Availability
National Guideline Clearinghouse	Comprehensive database of evidence-based clinical practice guidelines and related documents.	www.guidelines.gov	Free of charge
National Institute for Health and Clinical Excellence (NICE)	Organization responsible for providing UK guidance on promotion of good health. Includes evidence-based guidelines on maternal health.	www.nice.org.uk	Free of charge
Cochrane Library	International organization providing regularly updated evidence-based systematic reviews. Includes systematic reviews on maternal health.	www.cochrane.org	Free of charge
National Electronic Library for Health (NeLH)	Aimed at providing UK clinicians with the best current know-how and knowledge to support healthcare-related decisions. Includes a dedicated midwifery section.	www.nelh.nhs.uk	Free of charge
MIDIRS digest	High quality research papers from journals are selected by the midwife editor and abstracted.	Hardcopy digest	Subscription

Table 10.2 *Continued*

Resource	Description	Access	Availability
MIDIRS database searches	Standard and individual searches are available that scan the content of 500 English language journals, including midwifery, obstetrics and gynaecology, paediatrics, neonatal and key general medical and consumer titles.	www.midirs.org	Subscription
Bandolier	Evidence-based healthcare journal providing concise and easy to read summaries of the latest systematic reviews and meta-analyses. Includes maternal health section.	www.ebandolier.com	Free of charge
Clinical Evidence	Summarizes current knowledge and uncertainty about the prevention and treatment of clinical conditions, based on thorough searches and appraisal of the best available evidence. Includes section on maternal health.	www.clinicalevidence.com	Free of charge
Medline	The National Library of Medicine (US) online database of approximately 15 million biomedical and pharmaceutical citations, searched via natural language and/or controlled vocabulary.		Subscription
PubMed	The National Library of Medicine service, providing access to Medline back to the 1950s. Includes links to many full text articles and related resources.	www.ncbi.nlm.nih.gov/entrez	Free of charge
Embase	A powerful database providing access to several million biomedical citations, via natural language and/or controlled vocabulary searches.		Subscription
CINAHL	Database regarded as the authoritative source of information for the professional literature of nursing, allied health, biomedicine, and healthcare.		Subscription

PLANNING AN EFFICIENT SEARCH

Having devised a structured question it is possible to begin the search for evidence. Searching for evidence has been described as an essential clinical skill that demands to be efficient (Sackett et al 1997). With the unmanageable volume of information available to health professionals well documented, the need for efficiency is crucial (Greenhalgh 1997).

When planning a search it is important to consider the kind of evidence needed, and where it is most likely to be found. Research may be published or unpublished, can be found in non-peer-reviewed or peer-reviewed journals, and vary widely in quality (see Principles of good questions). Also, it is important to know how much time can be realistically spent, not only searching for but also appraising the evidence retrieved.

Midwives may elect to use resources that bypass the need to execute searches themselves. Beyond those previously described, there are also useful journals that provide structured abstracts as a means to alert clinicians to important advances, and up-to-date books, which may prove particularly useful for understanding physiology.

Many clinicians, however, are likely to execute searches by themselves using one or more biomedical database. The best well known are Medline, Embase, and the Cumulative Index to Nursing and Allied Health Literature, commonly known as CINAHL.

It is important to appreciate that confining a search to only one database risks the exclusion of either a majority, or large minority, of the available evidence. Research has demonstrated that using only one database to identify evidence is inadequate (Suarez-Almazor et al 2000). The overlap between Medline and Embase, in terms of journals indexed, is about 40% (Smith et al 1992), and studies have demonstrated that using both databases markedly improves the coverage of literature (Biarez et al 1991, Odaka et al 1992).

A considered search strategy is important when using resources such as Medline and Embase, each consisting of several million records. While Greenhalgh (1997) says that most people can learn to carry out a basic search of either database inside an hour, it is worthwhile investing some time in preparation and design of a search strategy before the searching begins in earnest. Straus et al (2005) provide a diagram of the steps for executing a pragmatic search for evidence (Fig. 10.1).

When using databases to search for evidence problems may be experienced. A search strategy may be highly sensitive, which will result in too much evidence, or a search may prove too specific, which will considerably lessen the return of evidence.

To counter such difficulties preparation is the key. Depending upon the results of a search it may prove necessary to narrow or widen the date range being searched; limit or expand the population sub-groups; determine the quality of evidence being sought; or limit evidence to that published in English. Most important of all, understand what it is that is actually being sought. Has a focused, well-structured question been asked? Has a population, comparison, intervention, outcome, and, if necessary, study design, been identified? Are the correct resources being employed that will most likely hold the evidence sought, and best utilize the available time?

The resources listed in Table 10.2 are examples, together with descriptions, that will greatly assist the search for evidence.

BASIC MEDLINE AND EMBASE SEARCH

A basic search for evidence using Medline or Embase can be undertaken in two ways:

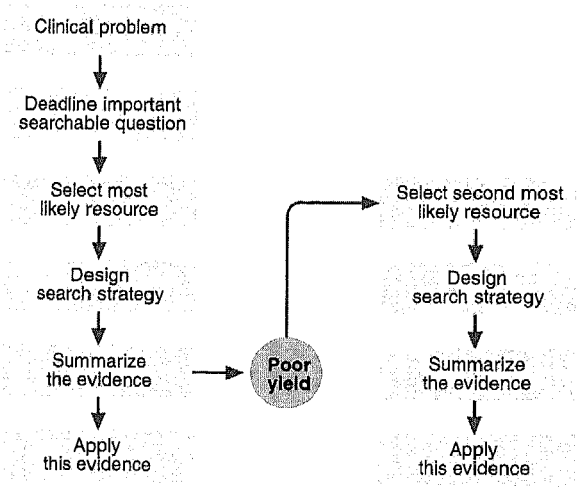

Figure. 10.1 General search strategy

- using *natural lang*uage (text words) to seek particular authors and institutions where research is carried out, and words in the title and/or abstract;
- using *Medical Subject Headings* (MeSH), or the controlled vocabulary of the databases, that constitute powerful theasuri.

An individual's ability to search for evidence will be greatly enhanced if they acquire an understanding of the following skills, adapted from Sackett et al (1997, p.73):

- the use of *both* natural language and MeSH;
- the identification of MeSH using the thesauri;
- the ability to identify *synonyms* directly related to MeSH and incorporate them into a search;
- the appropriate use of *search field tags* (e.g. abstract = .ab; author = .au; paper title word = .ti; textword = .tw);
- the ability to use search field tags as *limiters* (e.g. publication type = .pt; publication year = .py);
- the appropriate *truncation* of natural language and use of *wildcards* to replace characters within words;
- the ability to employ *adjacency commands* so linking words or phrases to each other;
- combining natural language, including search field tags, and MeSH by using *Booleon operators* (AND, OR, NOT) to expand and limit a search.

Medical librarians and information specialists have historically been regarded as experts in the field of searching for evidence; however, today there are many opportunities, via dedicated training courses and/or self-tuition, for health professionals to learn how to conduct effective and efficient searches.

ASSESSING AND INTERPRETING THE EVIDENCE

Once you consider that you have found the appropriate information to answer your question, it is necessary to assess the quality of the information to ensure that it is right for your purposes and that it provides valid evidence to answer your question. There are several accessible books that detail structured approaches to assessing research reports (see especially Greenhalgh 1997, Gray 1997, Sackett et al 1997, Straus et al 2005).

An essential first step in assessing the evidence is to discard poor-quality or irrelevant reports. Greenhalgh (1997) emphasizes the need to 'trash' papers and suggests that 'some purists would say 99% of published articles belong in the bin' (p. 34). She makes a strong argument that the quality of a paper is best assessed through the methods section and that the article should be 'trashed' on methods alone before looking at the results. She suggests three preliminary and basic questions as a way of getting an orientation to the paper are:

1. Why was the study done, what were the hypotheses, and what were the authors testing?
2. What type of study was carried out?
3. Was this research appropriate to the broad field of research studied?

Sackett et al (1997) propose the following questions to assess evidence:

- Is it true (valid)?
- Are the valid results important?
- Does it apply to the woman/women in my care?

VALIDITY (BOX 10.3)

When reviewing evidence from a study both internal and external validity (Box 10.3) of the results need to be considered. Internal validity is concerned with whether selection of groups and the way comparisons were carried out between the study's groups were sufficiently robust so that any reported difference is likely to be attributable to the effect being measured. External validity is concerned with the extent to which research findings can be generalized to people who are

Box 10.3

Validity: The degree to which the inference drawn from a study, especially generalizations extending beyond the study sample, are warranted when account is taken of the study methods, the representativeness of the study sample, and the nature of the population from which it is drawn.

Last 1995

similar to the participants but who did not take part in the study. A fundamental point when assessing evidence, and the validity of a study, is to question whether the methodology used is appropriate to the question posed (Box 10.4).

IMPORTANCE

Assessment of importance is related to the

- size and potential benefits of the effects measured in a study;
- probability of outcomes occurring over time;
- strength of association between the outcomes (either harmful or good) and interventions;
- increased probability of particular outcomes in different groups;
- precision of the estimates of effect.

Established methods are applied to numerical data reported in a study which are used to assess the importance of the results. For example, Straus et al (2005) describe the method to calculate how many people need to be treated to avoid an adverse outcome (number needed to treat; NNT) or to harm one person (number needed to harm; NNH). Measures of importance, strength of association and precision include relative risks, absolute risk reduction, Number needed to treat (NNT), and number needed to harm (NNH), odds, odds ratios and confidence intervals. The explanations of the terms by Straus et al (1998, pp. 141–145, 2005 pp.281–284) are outlined below:

- Odds – a ratio of non-events to events. If the event rate for a disease is 0.1 (10%), its non-event rate is 0.9 and therefore its odds are 9:1. Note this is not the same as the inverse of event rate.
- Odds ratio (OR) – is the odds of having the target disorder in the experimental group relative to the odds in favour of having the target disorder in the control group or the odds of being exposed in subjects with the target disorder divided by the odds in favour of being exposed in control subjects (without the target disorder).
- Risk ratio (RR) – is the ratio of risk in the treated group (EER) to the risk in the control group (CER) RR = ERR/CER.
- Relative risk reduction (RRR) – the proportional reduction in rates of bad outcomes between experimental and control groups.
- Absolute risk reduction (ARR) – the absolute arithmetic difference in rates of bad outcomes between experimental and control groups.
- Number needed to treat (NNT) – the number of patients who need to be treated to achieve one additional favourable outcome, calculated as 1/ARR and accompanied by a 95% confidence interval. If the ARR is 25% 1/25% = 4.
- Number Needed to Harm – the number of patients who need to be treated to achieve one additional unfavourable outcome, calculated as

Box 10.4 Broad topics of research

Most research studies are concerned with one or more of the following:

- Therapy – testing the efficacy of drug treatments, surgical procedures, alternative methods of service delivery, or other interventions. Preferred study design is randomized controlled trial.
- Diagnosis – demonstrating whether a new diagnostic test is valid (can we trust it?) and reliable (would we get the same results every time?). Preferred study design is cross sectional survey . . . in which both the new test and the gold standard test are performed
- Screening – demonstrating the value of tests that can be applied to large populations and that pick up disease at a presymptomatic stage. Preferred study design is cross sectional survey.
- Prognosis–determining what is likely to happen to someone whose disease is picked up at an early stage. Preferred study design is longitudinal cohort study.
- Causation–determining whether a putative harmful agent, such as environmental pollution, is related to the development of illness. Preferred study design is cohort or case-control study, depending on how rare the disease is . . . but case reports . . . may also provide crucial information.

Reproduced from Greenhalgh T 1997 How to read a paper. With kind permission from BMJ Books, BMJ Publishing Group.

1/ARR and accompanied by a 95% confidence interval. If the ARR is 25% 1/25% = 4.

- Confidence interval (CI) – expresses the range within which we would expect the true value of a statistical measure to fall. Few studies can be carried out amongst all of the people who would be eligible to experience the care or therapy being assessed. Therefore for a particular research study a sample is selected comprising people who it is hoped will be representative of the relevant population. This means that the results of that study should be considered as an estimate of measured effect(s), and need to be placed in the context of the likely upper and lower value of the effect should the same study be repeated with different representative samples. CIs are usually accompanied by a percentage value, which shows the level of confidence that we have that true value lies within this range. For example, for an NNT of 10 with a 95% CI of 5–15, we would have 95% confidence that the true value of NNT values was between 5–15.

APPLICATION TO WOMEN AND FAMILIES IN YOUR CARE

Questions of application are concerned with the similarity between the women and families in your care and those in reported research and should take into account their individual preferences and values.

APPRAISING EVIDENCE FOR VALIDITY AND IMPORTANCE

The assessment of the validity of particular studies will depend on the type of evidence used. Sackett et al (1997) propose the following categories:

- diagnosis
- prognosis
- harm
- therapy
- systematic reviews
- decision analysis
- qualitative research.

The following is based on the work of Straus et al (1998, 2005), using their categories for structured steps for appraisal. We have summarized many of the suggestions from three sources: Sackett et al (1997), Straus et al (1998, 2005) and the handbook from the 1998 Oxford workshop on teaching evidence-based medicine (University of Oxford 1998).

DIAGNOSIS

One of the most rapidly changing fields of maternity care lies in diagnosis and screening, particularly during the antenatal period. Midwives should be able to appraise current evidence and convey the accuracy of results to the women and families in their care. This area is perhaps one of the most complex of practice (see Chapter 11), and an understanding of key terms, such as sensitivity and specificity (see Box 10.5), is crucial not only to interpreting evidence, but also to interpreting test results and understanding their predictive value in practice.

Box 10.5 Diagnostic tests: key terms (reproduced from Sackett et al 1997, with permission)

Sensitivity: the proportion of people with the target disorder who have a positive test. It is used to assist in assessing and selecting a diagnostic test/sign/symptom.

SnNout: when a sign/test/symptom has a high sensitivity, a negative result rules out the diagnosis.

Specificity: the proportion of people without the target disorder who have a negative test. It is used to assist in assessing and selecting a diagnostic test/sign/symptom.

SpPin: when a sign/symptom has a high specificity, a positive result rules in the diagnosis.

Positive predictive value: the proportion of people with a positive test who have the target disorder.

Negative predictive value: the proportion of people with a negative test who are free of the target disorder.

Assessing validity

Straus et al (2005, pp 72–73) suggest using the following questions to appraise a paper or systematic review on diagnosis:

1. Was there an independent, blind comparison with a reference 'gold' standard of diagnosis?
2. Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?
3. Was the reference standard applied regardless of the diagnostic test result?

Was there an independent, blind comparison with a reference 'gold' standard of diagnosis?

Two criteria should have been met in order to answer this question in the affirmative. First, the women or babies in the study should have undergone both the diagnostic test in question (for example, nuchal fold scanning for Down syndrome) and the reference (or 'gold') standard of testing (e.g. amniocentesis for Down syndrome). Second, the person interpreting the tests of one should not know the results of the other, otherwise, consciously or subconsciously, the interpretation might be biased.

It is sometimes difficult to provide reference standards, but a definition of what is 'normal' has in some way to be agreed. Straus et al (2005, p. 69) use the diagnostic definition of normal alongside other definitions of normal, shown below in Box 10.6.

Was the diagnostic test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?

For example, if you are asking a question about the use of nuchal fold testing in all age groups of women, you would want it to be tested in a population of all age groups of women.

Was the reference standard applied regardless of the diagnostic test result?

Sackett says that when patients have a negative diagnostic test result, investigators are tempted to forego applying the reference standard, and when the latter is invasive (for example, an amniocentesis), it may be considered inappropriate to do so. Sackett suggests that if the report fails one or more of these criteria, you may wish to keep searching sources for further evidence.

Box 10.6 Six definitions of normal (reproduced from Sackett et al 1997, with permission)

1. Gaussian: the mean ± standard deviations. Assumes a normal distribution and means that all 'abnormalities' have the same frequency.
2. Percentile: within the range, say, of 5–95%. Has the same basic defect as the Gaussian definition.
3. Culturally desirable: preferred by society. Confuses the role of medicine.
4. Risk factor: carrying no additional risk of disease. Labels the outliers, who may not be helped.
5. Diagnostic: range of results beyond which target disorders become highly probable – the focus of this discussion.
6. Therapeutic: range of results beyond which the treatment does more good than harm. Means you have to keep up with advances in therapy.

Is this evidence about a diagnostic test important?

Once you have decided on the validity of the report or reports, it is appropriate to ask whether the evidence is important. Just because something is published does not mean that it is important!

Sackett et al (1997, p. 118) describe 'a modern way of thinking about diagnosis that takes into account both components of evidence-based medicine; your individual clinical expertise and the best external evidence'. This includes assessment of the prior assessment of possibilities before carrying out the test (prior or pretest probabilities) and the ability of the test to distinguish patients with and without the target disorder (sensitivity and specificity, and likelihood ratios). (See also Straus et al 2005, p. 82). The definitions of these terms are shown below.

Diagnostic tests: key terms

Sensitivity: the proportion of people with the target disorder who have a positive test. It is used to assist in assessing and selecting a diagnostic test/sign/symptom.

SnNout: when a sign/test/symptom has a high sensitivity, a negative result rules out the diagnosis.

Specificity: the proportion of people without the target disorder who have a negative test. It is used to assist in assessing and selecting a diagnostic test/sign/symptom.

SpPin: when a sign/symptom has a high specificity, a positive result rules in the diagnosis.

Positive predictive value: the proportion of people with a positive test who have the target disorder.

Negative predictive value: the proportion of people with a negative test who are free of the target disorder.

Likelihood ratio: the likelihood that a given test result would be expected in a patient with the target disorder compared with the likelihood that the same result would be expected in a patient without the target disorder.

Straus et al 2005, pp 282–283

Once you have determined validity and importance, the following questions help to determine whether or not you can apply this valid, important evidence to women and families in your care (Straus et al 2005, pp79–86):

- Is the diagnostic test available, affordable, accurate, and precise in your setting?
- Can you generate a clinically sensible estimate of your patient's pre-test probability (from practice data, from personal experience, from the report itself, or from clinical speculation)?
- Will the resulting post-test probabilities affect your management and help your patient? (Could it move you across a test-treatment threshold? Would your patient be a willing partner in carrying it out?)
- Would the consequences of the test help your patient?

PROGNOSIS/RISK ESTIMATION

Is this question about prognosis valid?

In general medicine the issue of prognosis:'How long have I got until I develop the condition in question' maps onto risk estimation in maternity care. An example question might be 'What is the increased probability of postpartum haemorrhage when a woman is gravida 6' (see Chapter 17). In assessing evidence related to prognosis for validity, we need to ask (adapted from Straus et al 2005, pp101–102):

1. Was a defined, representative sample of patients (women/fetuses/babies) assembled at a common (usually early) point in the course of their pregnancy?
2. Was follow up of participants sufficiently long and complete?
3. Were objective outcome criteria applied in a blind fashion?
4. If subgroups with different prognoses are identified
   - Was there adjustment for important prognostic factors?
   - Was there validation in an independent group of test set patients (women/fetuses/babies)?

The following example is designed to help think through whether the conclusions of a study about using maternal age as a prognostic factor for operative delivery were valid, and structured abstract (Rosenthal & Paterson-Brown 1998, p. 1064):

Objective To determine whether increasing maternal age increases the risk of operative delivery and to investigate whether such a trend is due to fetal or maternal factors.

Design Analysis of prospectively collected data on a maternity unit data base.

Population 6410 nulliparous women with singleton cephalic pregnancies delivering at term (37–42 weeks of gestation), nulliparous women with singleton cephalic pregnancies delivering at term between 1 January 1992 and 31 December 1995.

Setting The study was undertaken in a teaching hospital. There was a population of 6410. The results showed a positive highly significant association between increasing maternal age and obstetric intervention.

Main outcome measures Mode of delivery, rates of prelabour caesarean section, induction of labour and epidural usage.

Results There were a positive, highly significant association between increasing maternal age and obstetric intervention prelabour ($P < 0.001$) and emergency ($P < 0.001$) caesarean section, instrumental vaginal delivery (spontaneous labour $P < 0.001$; induced labour $P = 0.001$), induction of labour ($P < 0.001$). Epidural usage in induced labour and the incidence of small for gestational age newborns did not increase with increasing maternal age ($P = 0.68$ and $P = 0.50$, respectively).

Conclusions This study demonstrates that increasing age is associated with an incremental increase in obstetric intervention. Previous studies have demonstrated a significant effect in women older than 35 years of age, but these data show changes on a continuum from teenage years. This finding may reflect a progressive, age-related deterioration in myometrial function.

(Reproduced with kind permission from Blackwell Science Ltd.)

We can look at this paper against the framework of two of Sackett et al's (1997) questions:

Was a defined, representative sample of patients assembled at a common (usually early) point in the course of their disease?

This was a retrospective review of what is described as 'prospectively' gathered data which had been routinely collected. Information that might help to explain and define some of the measures (for example, the definition and diagnosis of fetal distress) was missing from the data collection. There are important gaps in the reported data, for example about the early pregnancy of women who participated in the study and about diagnosis of fetal distress. There are a number of indications in the report that the population studied may be different from the normal population of women giving birth. For example, the intervention rate in the hospital in which the study took place was described as higher than average and the women who were part of the study were described as having a higher than average age, and a larger proportion of 'career women'. This limits the confidence with which we might generalize the findings as the women in this study may not be representative of women in other populations. Length of follow-up was appropriate in a study that was concerned only with intervention rates.

Were objective criteria applied in a blind fashion?

Some of the criteria, for example operative delivery, are objective; others, such as failure to progress and fetal distress, are more subjective measures. There is no clear definition of criteria for the diagnosis of fetal distress and failure to progress (both of these being prone to bias in clinical interpretation). There is no mention of blinding for interpretation or for analysis. There was also no adjustment for prognostic factors. It is also possible that higher anxiety experienced by health professionals attending older women may be a factor in increasing the rate of interventions in older women, and as this study has not tested this important hypothesis. In studies examining the effect of maternal age on outcome that are adequately controlled for factors which may confound the outcomes, no significant differences are found (Harker & Thorpe 1992). There is no control or reanalysis for confounding factors, such as raised blood pressure, in this study.

The higher incidence of epidural analgesia in spontaneous labour may well have been an independent factor in increasing the intervention rate. Moreover, there is no mention of the fetal monitoring rate, which is likely to have been higher in this group of women, and which is in itself associated with a higher rate of intervention and a falsely high rate of diagnosis of fetal distress. A number of factors, including mobility and position in labour, may have affected the outcome; but none of these are described in the report. In addition, the lack of clear definition of failure to progress and fetal distress, which may be highly subjective assessments, is a fundamental flaw in this study. Attendant anxiety may well have affected the judgements made, in itself leading to a higher rate of intervention.

In assessing this study it is important to be aware of the idea that labelling particular women as being at increasing risk of intervention in labour becomes a self-fulfilling prophecy. It is quite likely that attendant anxiety or perception is likely to affect the outcome of labour (and this has not been refuted by this study). There is a danger of the outcomes of such studies being confounded, that

is, being affected by other factors than the one under study, in this case increasing age. Examples of potential confounders in this group of women would be the pre-existence of medical problems in subgroups of older women, or the tendency of health professionals to treat older women as being at higher risk, leading to a higher rate of continuous electronic fetal monitoring, which in itself produces a higher intervention rate. Unfortunately, the authors of this study did not heed the warnings of one of the papers they referenced (Harker & Thorpe 1992) and did not control for such confounding factors.

Is this evidence about prognosis important?

In relation to importance, two questions are relevant:

1. How probable are the outcomes over time?
2. How precise are the prognostic estimates?

Given the flaws of the study, its validity is in doubt and therefore it would not be appropriate to determine the importance.

Two further questions then follow.

Were the study patients similar to your own, and will this evidence make a clinically important impact on your conclusions about what to offer or tell your patients?

In this case, the flaws which have been identified in response to the questions above lead to the conclusion that the results are not valid. Therefore it would not be appropriate to change practice on the basis of the evidence from this study. Perhaps the only change in practice that should be considered would be to be more aware that bias caused by regarding older women as inherently at risk of adverse outcomes can itself increase the intervention rate.

HARM

Is this evidence about harm valid?

Straus et al (2005, p. 179) propose the following questions to assess the validity of studies to evaluate the possibility of harm:

1. Were there clearly defined groups of patients, similar in all-important ways other than exposure to the treatment or other cause?
2. Were treatment exposures and clinical outcomes measured the same ways in both groups (e.g. was the assessment of outcomes either objective (for example, death) or blinded to exposure)?
3. Was the follow up of study patients complete and long enough?
4. Do the results satisfy some 'diagnostic tests for causation'?
   - Is it clear that the exposure preceded the onset of the outcome?
   - Is there a dose response gradient?
   - Is there positive evidence from a challenge–rechallenge study?
   - Is the association consistent from study to study?
   - Does the association make biological sense?

This is a question of whether or not a treatment *caused* the harmful outcome experienced and thus, as Sackett et al (1997) tell us, 'benefits from what has been learned from classical epidemiology'. There are four possible designs for a study of the harmful effects of treatments. These are the randomized controlled trial, the cohort study, the case control study, or reports of one or two patients who have suffered from something that is unique and rare. Table 10.3 is a summary of the advantages and disadvantages of these approaches, as described by Sackett et al 1997.

Using evidence to inform practice

Let us look at the above questions with regard to a study to examine the effect of neonatal exposure to vitamin K on the risk of childhood cancer (Klebanoff et al 1993). The relationship between vitamin K and cancer was examined in a nested case control study that used data from the Collaborative Perinatal Project, a multicentre, prospective study of pregnancy, delivery and childhood that took place in the USA. Among 54,795 children born between 1959 and 1966, 48 cases of cancer were diagnosed after the first day of life and before the eighth birthday. Each case child was matched with randomly selected controls whose last study visit occurred at or after the age when the case child's cancer was diagnosed.

Table 10.3 Advantages and disadvantages of different kinds of harm study. (From Sackett et al 1997, with kind permission)

Type of study	Advantages	Disadvantages
Randomized controlled trial	Randomization would make groups similar for all other features that would cause harm	For rare events, very large trials would be needed (one per thousand would need 3000 patients in order to be 95% certain of seeing at least one adverse reaction)
Cohort study	Next most powerful design	The groups of patients (cohorts) may not be identical in every respect. Other things apart from the treatment being evaluated such as severity of illness may affect outcome. Same problem of size applies
Case control studies	For rare or late complications of treatment, need to rely on studies in which those who already have the disease are assembled and compared with a group who do not have the disease	The problem of confounding (of prognosis with exposure) is worse with case control studies because it may be impossible to measure confounders in a case even if they are known
Case reports and case series	Reports of one or two patients who developed a complication while under treatment (e.g. phocomelia in children born to women who took thalidomide)	May be enough but usually point to the need for further studies

Were there clearly defined groups of patients, similar in all-important ways other than exposure to the treatment or other cause?

The rarity of the disease in this case makes a randomized controlled trial impractical for answering the question of harm and makes the case control study the only practical design. Attempts were made to adjust for factors that might be associated with the development of childhood cancer (for example, race, sex, birth weight, maternal age, exposure to X-rays during pregnancy, and breast-feeding). Nevertheless, we cannot be confident that the groups were similar in every respect.

Were treatment exposures and clinical outcomes measured in the same ways in both groups (e.g. was the assessment of outcomes either objective, for example death, or blinded to exposure)?

The report states that 'two investigators who were blinded to the child's vitamin K status examined all records of children with cancer. Definite cases of cancer were required to have a histologically proved diagnosis of cancer, a clinical course including treatment consistent with the diagnosis or both' (Klebanoff et al 1993, p. 905).

Was the follow-up study of patients complete and long enough? There is no account of loss to follow-up, but the authors state that, afterwards, loss to follow-up was accounted for by life table methods. The study followed children up to 8 years of age.

Do the results satisfy some diagnostic tests for causation?

- *Is it clear that the exposure preceded the onset of the outcome?* It is as clear as can be that the exposure preceded the outcome. There was

reclassification for children with cancer before their first birthday, and therefore with the possibility that the cancer started in pregnancy.

- *Is there a dose–response gradient?* No dose–response gradient was available, but babies whose mothers were given vitamin K in the intrapartum period were excluded. All babies received the intramuscular vitamin K. In addition, there was an analysis of effect according to the brand of vitamin K used. There was then reanalysis to exclude the children in whom the administration of vitamin K was uncertain.

Because this was part of a larger study not aimed primarily at the evaluation of the effect of vitamin K, it is particularly important to be sure that vitamin K was actually administered when it was recorded and that all vitamin K was recorded when given. Recording in this situation was probably more careful because this was part of a research study with special documentation that was checked for completeness. In addition, there was recording by an observer in the delivery room of any drugs administered.

- *Is there positive evidence from a challenge–rechallenge study?* The challenge–rechallenge question (seeing what happens if a drug is withdrawn or re-administered) is not appropriate to this drug.
- *Is the result consistent from study to study?* The answer to this is no. The authors comment on two earlier studies including an evaluation of the effect of the administration of oral vitamin K that found twice the expected risk of cancer during childhood with the administration of vitamin K.
- *Does the association make biological sense?* There is no biological link made explicit in this study, but the fact that the incidence of childhood cancer has not increased with the frequent administration of vitamin K at birth increases confidence in the findings of the study.

Are the valid results from this harm study important?

Importance is evaluated against an estimation of the strength of the association between receiving the treatment and suffering the adverse effect. Strength here means the risk or odds of the adverse effect, with, as opposed to without, exposure to the treatment; the higher the risk or odds, the greater the strength and the more one should be impressed with it. Different tactics are used for estimating the strength of the association for different research methods. This is illustated in Table 10.4.

Using the data presented in the paper on vitamin K (Klebanoff et al 1993), this calculation is as shown in Table 10.5 and shows that, in this case, there is no association between Vitamin K and cancer.

Can the study results be extrapolated to your patient?

Given the situation in the USA, with such different demographic characteristics, direct extrapolation is not appropriate. One factor that is pointed out by the authors as being different is that the vehicle

Table 10.4 Calculating the strength of an association between a treatment and subsequent adverse outcomes. (Reproduced from Sackett et al 1997, with kind permission)

		Adverse outcome		Totals
		Present (Case)	Absent (Control)	
Exposed to the treatment	Yes (Cohort)	a	b	a+b
	No (Cohort)	c	d	c+d
	Totals	a+c	b+d	a+b+c+d

In a randomized trial or cohort study: relative risk = RR = [a/(a+b)]/[c/(c+d)] In a case-control study: relative odds = RO = ad/bc

Table 10.5 Calculating the strength of an association between vitamin K and cancer

		Adverse Outcome	
		Present (Case)	Absent (Control)
Exposed to the treatment	Yes vitamin K	33 a	171 b
	No vitamin K	c 15	d 69
	Controls	$\frac{33.69}{15.171}$	$\frac{2277}{2565} = 0.89$

of administration of the vitamin K differs between the USA and the UK.

THERAPY

If one wants to find out whether a treatment is likely to be of benefit, the most appropriate methodology to answer the question is a randomized controlled trial. There are so many factors that might influence the outcome of treatment that the only good way to control for possible sources of bias is to allocate people randomly to different treatment conditions.

In this section, we will look at assessing the evidence from a single study. If several studies are available, the investigator should first look for a systematic review and use that as a starting point.

Before going on to assess the results of a study, one must first ask the following questions.

Are the results of this single study valid?

The key questions to answer, following Straus et al (2005, p. 117)are:

1. Was the assignment of patients to treatments randomized?
2. And was the randomization list concealed?
3. Were patients and clinicians kept blind to which treatment was being received?
4. Aside from the experimental treatment, were the groups treated equally?
5. Were the groups similar at the start of the trial?

Studies of treatment (or therapy) compare outcomes in groups receiving the treatment with outcomes in groups either not receiving treatment or receiving an alternative treatment. It is essential that, at the outset, the groups being compared are as alike as possible. The only way to avoid bias when assigning people to groups is to make the assignments random. This does not guarantee that the groups will be identical, but it does ensure that any differences are more likely to be caused by chance alone.

For example, imagine that you are looking for a treatment for pregnancy-induced nausea and have heard that acupuncture worked. You decide to test this out by carrying out a randomized controlled trial to compare women who have been allocated acupuncture with those who have not. If you are testing the effectiveness of acupuncture compared to another form of care you would want to make sure that your groups consisted of a mixture of women with different severities of the condition, different lifestyles and a different tolerance of the symptoms: Uneven distribution of these factors between groups may confound your results by providing an alternative explanation for any result that you see, for example by exaggerating, counteracting or even cancelling out the effects of the therapy. If the random allocation is concealed from clinicians, they will be unaware of the treatment that the patient is receiving and will not be able to distort the effect, either consciously or unconsciously.

It would be important to set up a system so that you can ensure that you can account for all the women who were randomized, right through to the end of follow-up. This means that you would make every effort to account for women who were randomized but who dropped out from follow-up It is possible to analyse and report the results of a trial even when there has been 'loss' to follow-up by assigning those participants the 'worst' value of the outcome of interest. In the example of acupuncture for nausea in pregnancy, all the participants who did not contribute information to

follow-up might be considered to have experienced no improvement in nausea. However, as a rule of thumb, this approach is not useful if more than 20% of participants are lost to follow-up.

The correct method for a randomized controlled trial is to analyse results by 'intention-to-treat'. This means that the data are analysed according to the groups to which the participants were assigned when they were first randomized in the trial, and not according to their experience of care after assignment. For example, in a randomized controlled trial of home birth to assess the effect of home birth on intervention rates a number of women are likely to be transferred to hospital to give birth because of failure to progress or some other problem. If intention-to-treat analysis was not used to analyse the results, and outcomes are evaluated according to the location of birth, there would be a falsely low intervention rate in the home birth group. An intention to treat analysis provides an estimate of the effects of interventions in practice, by taking into account the reality that care and therapies do not always happen exactly as defined or prescribed.

In real practice it is often impossible to conceal (blind) the nature of allocated treatment/care for participants and clinicians. However it may be possible to have the outcomes being evaluated assessed by people who are unaware of what the original allocation was.

Are the results of this single preventive or therapeutic trial important?

Straus et al (2005, pp 126–130) propose calculating the number needed to treat (NNT), as a way of concealing how relevant and important the results of a study would be to the people you are working with. If a valid study found, for example, that you needed to treat four people to have one respond to the treatment, this gives a clear indication of how many people you would need to treat and the chance of a response in the population of people you are caring for. The view of these numbers differs for the midwife and the childbearing woman. A midwife may want to know that for every four women she treats, one will respond to treatment. For the childbearing woman, the most useful presentation of these figures is that she has a 1 in 4 chance of responding to therapy.

SYSTEMATIC REVIEW AND META-ANALYSIS

Although it is important to critically appraise the results of individual research studies, there are instances when it is possible to identify topics for which major work has already been achieved to synthesize and review evidence from several studies. In maternity care the Cochrane Library (www.cochrane org) is probably the most well-known important resource of reliable reviews of evidence. Increasingly other agencies are publishing the results of commissioned reviews to inform care (for example NICE May 2005, HTA 2005). Differences between systematic review and meta-analysis are outlined below.

A systematic review is a review that includes explicit and detailed description of why and how it was conducted so that it should be possible to replicate it (Jahad 1998). In should clearly state a research question and methods used to assemble data, explicitly taking into account issues such as bias, confounding and chance in interpretation of findings. Ideally its authors should seek to bring together research from all relevant evidence sources. It should clearly present the results of individual studies but should not combine results unless it is appropriate statistically to do so and issues such as bias and heterogeneity between studies have been satisfactorily dealt with.

Meta-analysis is a method of combining statistically the results of independent research studies, which are sufficiently similar, to generate a single estimate of effect for a particular treatment or therapy. The benefit to using this method is that it can increase the precision of the estimate of effect, thereby reducing uncertainty about what the range of the estimate is likely to be.

Straus et al (2005, p. 148) propose the following questions to test the validity of a systematic review:

1. Is it an overview of RCTs of the treatment you are interested in?
2. Does it include a methods section that describes:
   a. Finding and including all the relevant trials?

b. Assessing their individual validity?
3. Were the results consistent from study to study?

The first question asks whether you are sure that the treatment is the same as the one you are interested in and the others whether all the studies were carried out at the same, most powerful level of evidence.

For example, if a randomized controlled trial of vitamin K prophylaxis were available and you wanted to know whether oral vitamin K were effective, you would want to ask whether it was oral or IM vitamin K that had been tested.

Some overviews combine randomized and non-randomized studies. It is wrong to combine results from these different research methods and we would advise that you should not use results from such reviews to change practice.

DOES IT INCLUDE A METHODS SECTION THAT DESCRIBES (A) FINDING AND INCLUDING ALL THE RELEVANT TRIALS, AND (B) ASSESSING THEIR INDIVIDUAL VALIDITY?

The most important point to remember is that carrying out a systematic review is just like carrying out research. In other words, it uses the same approach as research to avoid bias and should be reported like research. As in the assessment of research, look carefully at the methods section. This should include a description of how the studies were identified which were included and excluded, and why. Straus et al (2005, p. 149) describe the importance of having sought unpublished results and of hand-searching for reports. The methods section should also say how the validity of the study was judged.

Were the results consistent from study to study?

Although we should not expect all trials to show exactly the same degree of effectiveness, it is reassuring if the results are not widely different.

Are the results of this systematic review important?

Deciding whether or not a treatment is important depends on the size and potential benefits of the effects of the treatment you are interested in.

Do these results apply to your patient?

- Is your patient so different from those in the overview that its results can't help you?
- How great would the potential benefit of therapy actually be for your individual patient (e.g. what is the NNT/NNH)?
- Do your patients and you have a clear assessment of their values and preferences?
- Are they met by this regimen and its consequences?

It is crucial to remember that systematic reviews and meta-analyses are no better than the studies they combine therefore applying critical appraisal methods to both is necessary.

ASSESSING QUALITATIVE RESEARCH

Qualitative research is important to all professionals working in the maternity services, but it is perhaps most important to midwives. Midwives hold the potential for strong and intimate relationships with childbearing women and their families as well as the potential for changing the experience of care and of pregnancy and birth.

Greenhalgh (1997, p. 151) describes clearly the limitations of quantitative research and the importance of qualitative research in 'seeking a deeper truth'. She quotes the aim of qualitative research as being 'to study things in their natural setting, attempting to make sense of, or interpret, phenomena in terms of the meanings that people bring to them', and that researchers use a 'holistic perspective which preserves the complexity of human behaviour (Denzin & Lincoln 1994)'. The contribution of social science and anthropology in researching maternity care is immense and provides a mine of information regarding the perspectives of childbearing women using the maternity services and their experiences, to inform midwives who want to understand better and improve care.

Greenhalgh (1997, see pp 155–61) suggests the following questions for assessing qualitative research. She is clear about the limits of such a checklist, and it should be used with caution. Qualitative research is by its very nature non-standard; and this list of questions is taken from

Greenhalgh to provide a structure. However, it is advisable to read Greenhalgh's chapter in full and to refer to the list of further reading at the end of that chapter.

DID THE PAPER DESCRIBE AN IMPORTANT CLINICAL PROBLEM EXAMINED THROUGH A CLEARLY FORMULATED QUESTION?

As with quantitative research, the topic area needs to be clearly defined. The process is iterative so the question may emerge more clearly at the end of the project, although it should still be clearly stated.

WAS A QUALITATIVE APPROACH APPROPRIATE?

It is most appropriate when the objective is to explore, interpret or obtain a deeper understanding of a particular issue.

HOW WERE THE SETTING AND SUBJECTS SELECTED?

The study should go beyond a convenience sample to a theoretical sample. Instead of taking an average view, the aim is to achieve an in-depth understanding of particular individuals, for example, a group of Somali women receiving maternity care in west London.

WHAT WAS THE RESEARCHER'S VIEW, AND HAS THIS BEEN TAKEN INTO ACCOUNT?

It is important to recognize that there is no way of controlling for observer bias in qualitative research. It is thus important that the investigator's personal perspective is fully explained.

WHAT METHODS DID THE RESEARCHER USE FOR COLLECTING DATA, AND ARE THESE DESCRIBED IN ENOUGH DETAIL?

The methods section is likely to be lengthy and discursive. You should ask the question. 'Have I been given enough information on methods?' There are no hard and fast rules.

WHAT METHODS DID THE RESEARCHER USE TO ANALYSE THE DATA, AND WHAT QUALITY CONTROL MEASURES WERE IMPLEMENTED?

The researcher must have found a systematic way of analysing the data. A number of methods are available, which include content analysis. A good paper will describe a method of quality control; in other words, it will not depend on just the interpretation of one person.

ARE THE RESULTS CREDIBLE?

The results should be independently and objectively verifiable. Are they sensible and believable, and do they matter in practice?

WHAT CONCLUSIONS WERE DRAWN, AND WERE THEY JUSTIFIED BY THE RESULTS?

Greenhalgh suggests using these three questions:

- How well does this analysis explain why people behave in the way they do?
- How comprehensive would this explanation be to a thoughtful participant in the setting?
- How well does the explanation cohere with what we already know? (Mays & Pope 1996).

CONCLUSION

It is easy to feel overwhelmed by the never-ending deluge of information about developments and changes in practice and care. The world-wide web has given everyone with access the possibility of finding almost infinite sources of information and sources of evidence that healthcare practitioners use are increasingly available to the women and families using maternity services. This should lead to invigorating information-sharing and into exciting and innovative realms of practice. However it must not mean abdication of responsibility by midwives for retrieving and appraising evidence, and thinking through its implications to midwifery practice and to the wellbeing and health of women and families. In this chapter we outlined structured and systematic approaches to

use to find, assess and evaluate evidence. Just as you would apply a structure when undertaking a clinical examination, this framework will help you to make sure that you have covered key issues and think through whether there is anything important that has been missed out.

It is important not to be intimidated by the mystique and sense of élitism that sometimes surrounds research in midwifery. Do not be afraid to make judgements about whether or not research findings make sense on a basic level. Your everyday experience and the knowledge you gain from practice give you an important basis for judging whether or not things simply make sense. Although you should be aware of the limitations of personal knowledge in making generalizations, your general sense about whether or not a piece of research is 'somehow just not right' can be an invaluable tool to cut through the information 'jungle'.

We believe that asking questions that arise from practice and being able to find, evaluate, and implement the findings of research are crucial to ensuring that midwifery care is likely to be beneficial and to avoid harm. However, we often find that there is no strong evidence to help answer questions arising from practice. In that case, the challenge is to be honest about the sources we use for answering questions and dealing with uncertainty. For us, and we think for midwifery practice generally, the guiding principals of effective evidence-based midwifery are to make use of the best evidence possible to underpin ethical maternity care for all women and their families. And always to be prepared to challenge ourselves – as much as others!

POINTERS FOR PRACTICE

- Above all, evidence-based practice demands that practitioners constantly question whether or not they are doing good or harm.
- Clear and well-constructed questions are fundamental for evidence-based care.
- Systematic and planned approaches are necessary to find and assess evidence.
- The outcomes of maternity care that are of concern to midwives and women and families are broader than just measures of mortality and physical morbidity and encompass such issues as psychosocial wellbeing, personal and family safety, economic outcomes.
- There are a number of specialist resources for midwives (for example MIDIRS), and these should be available and used whenever possible.

References

Biarez O, Sarrut B, Doreau C G, Etienne J 1991 Comparison and evaluation of nine bibliographic databases concerning adverse drug reactions. Drug Intelligence and Clinical Pharmacy 25: 1062–1065

Chalmers I, Enkin M E, Kierse M J N C 1989 Effective care in pregnancy and childbirth. Oxford University Press, Oxford

Cochrane A 1972 Effectiveness and efficiency; random reflections on health services. The Nuffield Provincial Hospitals Trust, London

Denzin N K, Lincoln Y S 1994 Handbook of qualitative research. Sage, London

Gray J A Muir 1997 Evidence-based healthcare: how to make health policy and management decisions. Churchill Livingstone, Edinburgh

Greenhalgh T 1997 How to read a paper: the basics of evidence-based medicine. BMJ Publishing Group, London

Harker L H, Thorpe K 1992 'The last egg in the basket'? Elderly primiparity – a review of findings. Birth 19(1): 23–28

HTA Online. Available: http://www.hta.nhsweb.nhs.uk/projectdata/3_project_record_published.asp?PjtId=1011 (accessed 8th August 2005)

Jahad J 1998 Randomised controlled trials. BMJ, London

Klebanoff M A, Read J S, Mills J L, Shiono P H 1993 The risk of childhood cancer after neonatal exposure to vitamin K. New England Journal of Medicine 329(13): 905–908

Last J M 1995 A dictionary of epidemiology. 3rd edn. Oxford University Press, Oxford

Mays N, Pope C 1996 Qualitative research in healthcare. BMJ Publishing Group, London

NICE May 2005 Breastfeeding for longer – what works. Systematic review summary. Online.

Available: http://www.publichealth.nice.org.uk/page.aspx?o=511618 accessed October 19th 2005

Odaka T, Nakayana A, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T et al 1992 The effect of a multiple literature database search. A numerical evaluation in the domain of Japanese life science. Journal of Medical Systems 16(4): 177–181

Olsen O 1997 Meta analysis of the safety of home birth. Birth 24: 4–13

Page L 2000 Using evidence to inform practice. In: Page L (ed.) The new midwifery. Science and sensitivity in practice. Churchill Livingstone, London, p. 45

Rosenberg W, Donald A 1995 Evidence-based medicine: an approach to clinical problem-solving. British Medical Journal 310: 1122–1125

Rosenthal A, Paterson-Brown S 1998 Is there an incremental risk of obstetric intervention with increasing maternal age? British Journal of Obstetrics and Gynaecology 105: 1064–1069

Sackett D L, Richardson W S, Rosenberg W, Haynes B R 1997 Evidence-based medicine: how to practise and teach EBM. Churchill Livingstone, Edinburgh

Silverman W A 1980 Retrolental fibroplasia: a modern parable. Academic Press, London.

Silverman WA 1998 Where's the evidence. Debates in modern medicine. OUP, Oxford

Smith B, Darzins P, Quinn M, Heller R F 1992 Modern methods of searching the medical literature. Medical Journal of Australia 157: 603–611

Straus E S, Badenoch D, Scott Richardson W, Rosenberge W, Sackett D L 1998 Practising evidence-based medicine. Tutor's manual. 3rd edn. Radcliffe Medical, Oxford

Straus E S, Scott Richardson W, Glasziou P, Haynes R B 2005 Evidence based medicine: how to teach and practice EBM. 3rd edn. Elsevier, Edinburgh

Suarez-Almazor M E, Belseck E, Homik J, Dorgan M, Ramos-Remus C 2000 Identifying clinical trials in the medical literature with electronic databases: Medline alone is not enough. Controlled Clinical Trials 21: 476–487

University of Oxford 1998 Manual of Oxford workshop on teaching evidence-based medicine. University of Oxford, Oxford

USEFUL ADDRESSES

National Institute for Health and Clinical Excellence
Mid City Place
71 High Holborn
London
WC1V 6NA
www.nice.org.uk

MIDIRS
9 Elmdale Road
Clifton
Bristol
BS8 1SL
www.midirs.org